Extremal Human Curves: a New Human Body Shape and Pose Descriptor

Rim Slama, Hazem Wannous, Mohamed Daoudi

Extremal Human Curves: a New Human Body Shape and Pose Descriptor

Rim Slama, Hazem Wannous and Mohamed Daoudi

***Abstract*— Automatic estimation of 3D shape similarity from video is a very important factor for human action analysis, but also a challenging task due to variations in body topology and the high dimensionality of the pose configuration space. We consider the problem of 3D shape similarity in 3D video sequence for different actors and motions. Most current approaches use conventional global features as a shape descriptor and define the shape similarity using L_2 distance. However, such methods are limited to coarse representation and do not sufficiently reflect the pose similarity of human perception. In this paper, we present a novel 3D human pose descriptor called Extremal Human Curves (EHC), extracted from both the spatial and the topological dimensions of body surface. To compare tow shapes, we use an elastic metric in Shape Space between their descriptors, based on static features, and then perform temporal convolutions, thereby capturing the pose information encoded in multiple adjacent frames.**

We quantitatively analyze the effectiveness of our descriptors for both 3D shape similarity in video and content-based pose retrieval for static shape, and show that each one can contribute, sometimes substantially, to more reliable human shape and pose analysis. Experimental results are promising and show the robustness and accuracy of the proposed approach by comparing the recognition performance against several state-of-the-art methods.

I. INTRODUCTION

While human analysis in 2D image and video has received great interest during the last two decades, 3D human body is still a little explored field. Relatively few authors have so far reported work on 3D static analysis of 3D human body, but still less on 3D human video analysis.

3D Human body shape similarity is itself an important area, recently attracted much attention in the field of human-computer interface (HCI) and computer graphics, with many related research studies. Among these, researches started with 3D features have been applied for body pose estimation and 3D video analysis. More than that, 3D video sequences of human motion is more and more available. In fact, their acquisition with a multiple view reconstruction systems or animation and synthesis approaches [1] [2] received a considerable interest over the past decade following the pioneering work of kanade [3].

Several potential applications arisen from this, such as content based pose retrieval in a basis of human, applications of motion transition decision and concatenating 3D video sequences to produce a novel character animation, 3D video summarization and compression and 3D mesh video retrieval. These potential applications subsequently require solving the problem of identifying frames with similar pose.

In this paper, we present a novel 3D human curve-based shape descriptor called Extremal Human Curve (EHC) descriptor, extracted from body surface; robust to topology changes and invariant to rotation and scale. It is based on extremal features and geodesics between each pair of them. Every 3D frame will be represented by a collection of open curves whose comparison will be performed in a Riemannien Shape Space using an elastic metric. Our ultimate goal is to be able to perform reliable reduced representation based-geodesic curves for shape and pose similarity metric, which can be employed in several potential applications like video annotation and concatenation, activity analysis and behaviour understanding.

Our method contributes to the challenges of designing effective shape similarity tool in three ways. First, it uses an efficient descriptor that can be applied to many existing 3D shape retrieval tools. Second, the EHC provides a reduction in the dimensionality of the shape representation, thereby reducing both the space for storage and the time for comparison. Finally, the elastic metric applied in the Riemannien Shape Space gives more efficient similarity distance than those obtained by rotation invariant approaches, e.g. Shape Histogram and Spherical Harmonic.

The rest of the paper is structured as follows. The next section discusses previous works in the area of shape description, similarity metrics and properties. The extremal curves extraction and curves collection on the human body surface are presented in section III. In section IV, we describe the mathematical properties of the elastic metric in the Shape Space. In section V, we provide experimental tests comparing 3D shape similarity metric results over different databases with others approaches of the state-of-the-art. Finally, we conclude in section VI by summarizing our results and discussing issues for future work.

II. RELATED WORK

In this section, we review some methods of shape and pose similarity, related to our approach, which only utilize the full-reconstructed 3D data for feature extraction and description. A number of researchers have been addressed the problem of shape similarity for 3D video. Most of them evaluate a similarity metric on spatial shape descriptors based on surface or volume. Johnson and Hebert proposed the

This work is supported by the Region Nord-Pas-de-Calais and by the French Delegation for Research and Technology.

Rim Slama is with the LIFL Lab., Lille University/Telecom Lille1, France. rim.slama@telecom-lille1.eu

Hazem Wannous is with the LIFL Lab., Lille University/Telecom Lille1, France. hazem.wannous@telecom-lille1.eu

Mohamed Daoudi is with the LIFL Lab., Institut Mines-Telecom/Telecom Lille1, France. mohamed.daoudi@telecom-lille1.eu

spin image [4], encoding the density of mesh vertices into 2D histogram. Osada et al. use a Shape Distribution [5] to measure geometric properties of a 3D model, by computing the distance between random points on the surface, and then construct its shape signature. Ankerst et al. introduced the shape histogram [6], as a volume sampling spherical histogram by partitioning the space containing an object into disjoint cells corresponding to the bins of the histogram. Kazhdan et al. applied spherical harmonics to describes an object by a set of spherical basis functions [7]. These approaches use global features to characterize the overall shape and provide a coarse description, that is insufficient to distinguish similarity in 3D video sequence for an object having the same global properties in the time.

Some other works have trends to capture the evolvement of shape and pose changes in the sequence and then to add temporal information [8]. The temporal similarity in 3D video has also been addressed in the case of skeletal motion [9]. They demonstrate that skeleton-based Reeb-Graph have excellent performance in the task of finding similar poses of the same person in 3D video, and has recently presented as a stable topology descriptor, preserving the geometrical representation in presence of deformations [10].

Recently, Huang et al. [11] proposed 3D shape similarity metrics for 3D video sequences of people, using time filtering and shape flows obtained via invariant-rotation shape histogram. Their experiments conducted for 3D video sequence from i3DPost database [12] showed that Shape Histogram performs all other descriptors and gives the best recognition performance for time-varying non-rigid shape retrieval. Such approaches give a good shape descriptor but usually do not capture any geometrical information about the 3D human body pose and joint positions / orientations. This prevents its use in certain applications that require accurate estimation of the pose of the body parts. This approach will be discussed in this paper by comparing their results to the ours, both obtained over i3DPost database.

III. Extremal Curves

We aim to present a body shape as a skeleton based shape representation. This skeleton will be extracted on the surface of the mesh by connecting features located on the extremities of the body. The main idea behind the use of this representation is to analyze pose variation with elastic deformation of the body, using representative curves on the surface.

A. Feature point detection

Feature points refers to the points of a surface located at the extremity of its prominent components. In our approach, theses feature points are used to present a new pose descriptor based on curves connecting each two extremities. To extract the feature points, some works use Gaussian curvature threshold [13] or multidimensional scaling [14] and others, more robust, propose a cross-analysis using geodesic based scalar functions defined over the surface [15]. Feature points in this later are points resulting from the intersection of their two sets of local extrema. We chose to detect the body extremities by this method since it is based on geodesic distance evaluation, stable and invariant to geometrical transformations and model pose. Fig. 1 (a) shows the stability of the feature extraction for different persons (shape) in different poses.

B. Body curves extraction

Let M be a body surface and $E = \{e_1, e_2, e_3, e_4, e_5\}$ a set of feature points on the body representing the output of feature point extraction. Let β denote the open curve on M which joints two feature points of M $\{e_i, e_j\}$. To obtain β, we seek for geodesic path P_{ij} between e_i and e_j. We repeat this step to extract extremal curves from the body surface ten times so that we do all possible paths between elements of E. As illustrated in Fig.1 (b) the body is represented using these extremal curves $M \sim \bigcup \beta_{ij}$.
We have chosen to represent the body pose by a collection of curves for two reasons. Firstly, these curves connect limbs and give obviously a good representation of the body shape and pose, using a reduced representation of the mesh surface. Secondly, elastic analysis shapes of curves inside Shape Space is more efficient [16]. However, to compare correspondent extremal curves we need a distance to evaluate how much the shape of the corresponding curves is similar. The distance we are going to use is called an elastic metric.

IV. Elastic metrics in Shape Space

While human body is an elastic shape, its surface can be simply affected by a stretch (raising hand) or a bind (squatting). In order to analyze human curves independently to this elasticity, we need an elastic metric within a Shape Space framework [17].

A. Elastic distance

Let $\beta : I \to \mathbb{R}^3$, for $I = [0, 1]$, represents an extremal curve obtained as described above. To analyze the shape of β, we shall represent it mathematically using a *square-root velocity function* (SRVF), denoted by $q(t)$:

$$q(t) \doteq \frac{\dot{\beta}(t)}{\sqrt{\|\dot{\beta}(t)\|}} \,. \tag{1}$$

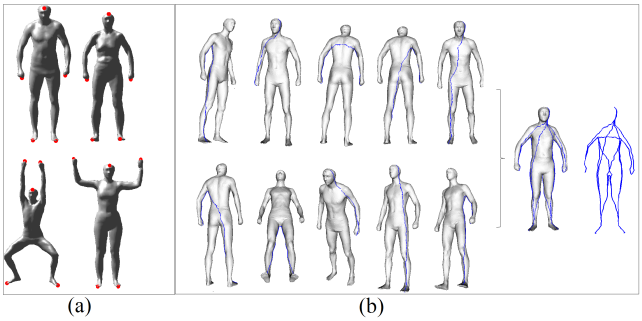

Fig. 1: Body curves extraction: (a) feature points extracted from human body surface, (b) human body represented as a collection of extremal curves.

$q(t)$ is a special function introduced by [16] that captures the shape of β and is particularly convenient for shape analysis. It has been shown in [16] that the classical elastic metric for comparing shapes of curves becomes the $\mathbb{L}^2$-metric under the SRVF representation. This point is very important as it simplifies the calculus of elastic metric to the well-known calculus of functional analysis under the $\mathbb{L}^2$-metric. We define the set:

$$\mathcal{C} = \{q : I \to \mathbb{R}^3 | \|q\| = 1\} \subset \mathbb{L}^2(I, \mathbb{R}^3) \ . \tag{2}$$

With the $\mathbb{L}^2$ metric on its tangent spaces, $\mathcal{C}$ becomes a Riemannian manifold. In particular, since the elements of $\mathcal{C}$ have a unit $\mathbb{L}^2$ norm, $\mathcal{C}$ is a hyper sphere in the Hilbert space $\mathbb{L}^2(I, \mathbb{R}^3)$. In order to compare the shapes of two extremal curves, we can compute the distance between them in $\mathcal{C}$ under the chosen metric. This distance is defined to be the length of the shortest geodesic connecting the two points in $\mathcal{C}$. Since $\mathcal{C}$ is a sphere, the formulas for the geodesic and the geodesic length are already well known. The geodesic length between any two points $q_1, q_2 \in \mathcal{C}$ is given by:

$$d_c(q_1, q_2) = cos^{-1}(\langle q_1, q_2 \rangle) \ , \tag{3}$$

and the geodesic path $\alpha : [0, 1] \to \mathcal{C}$, is given by:

$$\alpha(\tau) = \frac{1}{\sin(\theta)} \left(\sin((1-\tau)\theta)q_1 + \sin(\theta\tau)q_2\right) \ ,$$

where $\theta = d_c(q_1, q_2)$ is the inner product in the Hilbert space L2.

It is easy to see that several elements of $\mathcal{C}$ can represent curves with the same shape. For example, if we rotate a body changing its direction in $\mathbb{R}^3$, and thus its extremal curves, we get different SRVFs for the curves but their shapes remain unchanged. Another similar situation arises when a curve is re-parametrized; a re-parametrization changes the SRVF of curve but not its shape. In order to handle this variability, we define orbits of the rotation group $SO(3)$ and the re-parametrization group Γ as equivalence classes in $\mathcal{C}$. Here, Γ is the set of all orientation-preserving diffeomorphisms of I to itself and the elements of Γ are viewed as re-parametrization functions. For example, for a curve $\beta : I \to \mathbb{R}^3$ and a function $\gamma \in \Gamma$, the curve $\beta \circ \gamma$ is a re-parametrization of β. The corresponding SRVF changes according to $q(t) \mapsto \sqrt{\dot{\gamma}(t)}q(\gamma(t))$. We define the equivalent class containing q as:

$$[q] = \{\sqrt{\dot{\gamma}(t)}Oq(\gamma(t)) | O \in SO(3), \quad \gamma \in \Gamma\} \ ,$$

The set of such equivalence class is called the shape space $\mathcal{S}$ of elastic curves [16].

Let $q_2^*(t) = \sqrt{\dot{\gamma^*}(t)}O^*q_2(\gamma^*(t)))$ be the optimal element of $[q_2]$, associated with the optimal rotation O^* and re-parametrization γ^* of the second curve, then

$$d_s([q_1], [q_2]) \doteq d_c(q_1, q_2^*) \ , \tag{4}$$

In practice, SVD is used to compute optimal rotation and the dynamic programming is performed for optimal parametrization.

The shortest geodesic between $[q_1]$ and $[q_2]$ in $\mathcal{S}$ is given by:

$$\alpha(\tau) = \frac{1}{\sin(\theta)} \left(\sin((1-\tau)\theta)q_1 + \sin(\theta\tau)q_2^*\right) \ ,$$

where θ is now $d_s([q_1], [q_2])$.

B. Static shape similarity

The elastic metric applied on extremal curve-based descriptors can be used to define a similarity measure. Given two 3D meshes x, y and their descriptors $x' = \{q_1^x, q_2^x, q_3^x, ..., q_N^x\}$ and $y' = \{q_1^y, q_2^y, q_3^y, ..., q_N^y\}$, the mesh-to-mesh similarity can be represented by the curve pairwise distances and can be defined as follows:

$$s(x, y) = d(x', y') \ , \tag{5}$$

$$d(x', y') = \frac{\sum_{i=1}^{N} d(\beta_i^x, \beta_i^y)}{N} = \frac{\sum_{i=1}^{N} d_s([q_i^x], [q_i^y])}{N}. \tag{6}$$

where N is the number of curves used to describe the mesh. The mean of curve distances between two descriptors captures the similarity between their mesh poses. In case of change of shape in even one curve, the global distance will be affected and increase indicating that the poses are different.

In Fig.2, a geodesic path between each corresponding two extremal curves, taken from two human bodies doing different poses, is computed in Shape Space. For the left model, the person's arm is down and for the right model it is raised. In the middle the geodesic path between each two curves is shown in the shape space. This evolution looks very natural under the elastic matching. Since we have geodesic paths denoting optimal deformations between individual curves, we can combine these deformations to obtain full deformations between two poses. In order to have a global distance, an arithmetic distance is computed. Thanks to this global distance, we can compare human poses. For small deformation, the distance will be small and it is going to increase for models doing different poses.

V. Experimental results

To show the practical relevance of our method, we perform an experimental evaluations on several databases, and compare it, separately, to the most efficient descriptors of the state-of-the-art methods. We firstly evaluate our descriptor for content-based pose retrieval application over public static shape database and evaluate the results against Spherical Harmonic descriptor [7]. Secondly, we measure the efficacy of our descriptor to capture the shape similarity in 3D video sequences of different actors and actions from another public

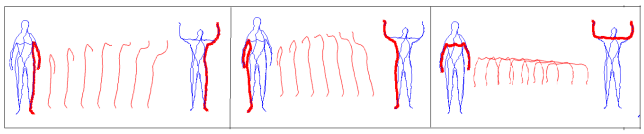

Fig. 2: Geodesic path between three extremal human curves of a neutral pose with raised hands.

database. We evaluate this later against Temporal Shape Histogram [11], Multi-resolution Reeb-graph [9] and other classic shape descriptors, using provided Ground Truth.

A. Content-based pose retrieval

The crucial point in all content-based retrieval systems is the notion of "similarity" employed to compare different objects. In fact, thanks to the static shape similarity, we are able to compare human poses using their extremal human curve descriptors and decide if two poses are similar or not. In this experiment, we advocate the usage of the EHC for content-based pose retrieval, where a query consists of a 3D human shape model in a given pose. As in a classical retrieval procedure, in response to a given 3D shape query, our approach searches the benchmark database and returns an ordered list of responses called the ranked list. The evaluation of the algorithm is then transformed to the evaluation of the quality of the ranked list.

The similarity metric represented by elastic measure values between each pair of models allows us to generate a confusion matrix for all classes of pose, in order to evaluate the recognition performance by computing statistic retrieval measures thanks to the provided ground truth. Once extremal curves are extracted from all models of the database and all query-to-model distances for a given query are calculated, we have to analyze the efficiency of distances based-curve.

1) Curve selection: From five feature endpoints, we have extracted ten extremal curves representing the human body shape model. According to the human poses, extremal curves exhibit different performance and some curves are more efficient to capture the shape similarity between two poses. Our shape descriptor can be seen as a concatenation of ten curve representations and the similarity between two shape models doing two different poses, is represented by a vector of ten elastic distance values. Before all tests, we analyze the performance of all possible combinations of curves on the shape similarity measurements. A Sequential Forward Selection method, applied on elastic distance values and coupled with First-tier criterion, has been used to select the best combination of curves among all possible ones (1013 combinations).

This experiment has been evaluated on a set of shape models of different persons from a statistical shape database representing different poses with known ground truth [18]. Empirical tests show that best combination is obtained by the five curves: right hand to right foot, left hand to left foot, left hand to right hand, left foot to right foot, and head to the right foot. The selected five curves seem to be the most stable ones and are sufficient to represent at best the body like a skeleton on the surface. Therefore, the elimination of five curves allows to eliminate the ambiguity due to the redundancy of some curves of the body parts.

2) Results from static shape data: To assess the performance of the EHC for content-based pose retrieval, several experiments were performed on a statistical shape database [18]. This database is challenging for human body shapes and pose retrieval as it is realistic shape database captured with a 3D laser scanner, and interesting as it contains more than hundred subjects doing more than thirty different poses. We perform our descriptor on a subset of 338 shape models obtained from different subjects with 18 consistent poses (p0, p1, p2, p3, p4, p5, p6, p7, p8, p9, p10, p11, p12, p13, p16, p28, p29, p32) [18]. Each pose represents a class where at least 4 different subjects do the same pose.

The self similarity matrix obtained from the mean elastic distance of the five selected curves is shown in the Fig.3. Main observation made from this matrix is that similar poses have a small distance which increases with the degree of the change between poses. This allows pose classification or pose retrieval by comparing models using their extremal curve representation and the elastic metric for measuring the distance.

We compare our descriptor to the popular Spherical Harmonic descriptors [7], applied with 32 shells, 16 descriptors for each shell and Euclidean distance as similarity measurement. In Fig.4 Recall/Precession curves show very good results due to the efficient of extremal curves to detect the information of the pose. Our EHC, using the five selected curves, outperforms SH descriptor to retrieve models with the same pose. Notice that the accuracies for very little number of poses are relatively low. This is probably due to the significant variations of the same pose performed by different subjects. For example, some ambiguities can be noticed in the case of a pose where the body just twist the torso due to the confusion with a neutral pose. Heavy move of legs can also be not detected and confused with the neutral pose.

B. Shape Similarity for 3D Video Sequences

Identifying frames with similar shape and pose can be used potentially for concatenative human motion synthesis. Concatenate existing 3D video sequences allows the construction of a novel character animation. It can also be used for the extraction of key-frames for video summarization, by analyzing self-similarity matrix. Similar frames will be grouped

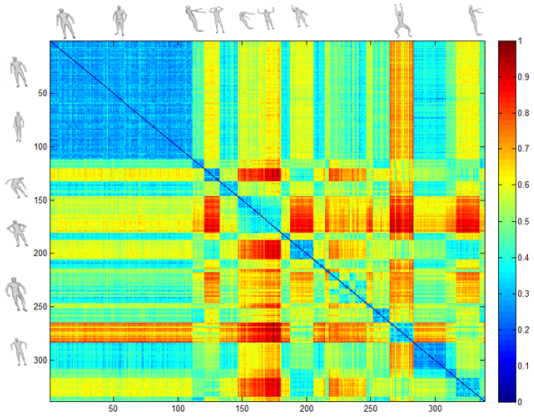

Fig. 3: Confusion similarity matrix. The matrix contains pose similarity computation between models of a 3D humans in different poses. The blocks with coldest color allow to identify models having similar poses.

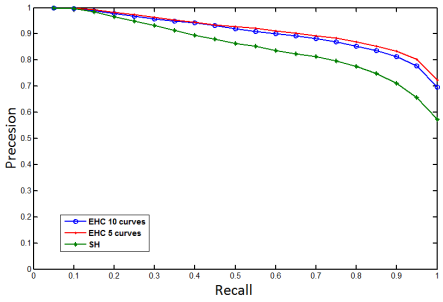

Fig. 4: Precision/Recall curves for EHC and SH.

and representative ones will be selected to summarize the video. A good descriptor that matches correctly correspondent frames allows the synthesis of videos with smooth transitions and finding best frames to summarize the video. To evaluate the effectiveness of our descriptor to correctly identify similar frames, we perform an experiment using a ground-truth database from synthetic 3D video sequences of people, and compare the recognition performance against several state-of-the-art descriptors. To compute similarity measure with consideration of multiple adjacent frames, we extend the static shape similarity (EHC) to a temporal one (TEHC) using a simple time filtering convolution.

1) Similarity metric and evaluation criterion: Given two 3D video sequences $P = \{p_i\}$ and $Q = \{q_j\}$, we firstly represent each pose frame by an EHC and compute the frame-to-frame similarity between p_i and q_j as $s_{ij} = s(p_i, q_j)$. Then, in order to evaluate temporal similarity according to [11], we apply a simple time filter with a window size $2N_t + 1$. This time filter is a way of incorporating motion in the similarity measure, and temporal similarity is computed as follow [11],

$$s_{ij}^t = \frac{1}{2N_t + 1} \sum_{k=-N_t}^{N_t} s(i + k, j + k) \tag{7}$$

A quantitative evaluation have been conducted over a syntectic database [12], created using 14 models, with different body-shape and clothing, animated using 28 motion capture sequences. Each sequence comprised 100 frames giving a total of 39200 frames of synthetic 3D video with known ground-truth correspondence. A Temporal ground truth similarity between two frames is defined as a combination of shape and velocity similarity as described in [11].

In order to identify frames as similar or dissimilar, a threshold is set on temporal ground truth similarity matrix. Recognition performance is evaluated using the Receiver-Operator-Characteristic (ROC) curves, created by plotting the fraction of true-positive rate (TPR) against the fraction of false-positive rate (FPR), at various threshold settings. The true and false dissimilarity compare the predicted similarity between two frames, against the ground-truth. An example of self-similarity matrix computed using ground-truth descriptor, static and temporal descriptors is shown in Fig.5. This figure illustrates also the effect of time filtering with increasing temporal window size for EHC descriptors on a periodic walking motion.

2) Comparison with state-of-the-art descriptors: A comparison is made between our TEHC (Temporal Extremal Human Curve) and several descriptors from the state-of-the-art: Shape Distribution (SD) [5], Spin Image (SI) [4], Spherical Harmonics Representation (SHR) [7], two Shape-flow descriptors, the global / local frame alignment Shape Histograms (SHvrG / SHvrS) [11] and Reeb-Graph as skeleton based shape descriptors (aMRG) [19] [9]. Note that a spectral representation was also evaluated in [9] which is the Multi-Dimentional Scaling (MDS).

To measure the performance of the similarity metric results, we plot the ROC curves obtained from our EHC descriptor (see Fig.6 (a)). These results are compared with ROC curves obtained by all state-of-the-art descriptors presented in figure 6 at [9] where our descriptor is among the more three efficient descriptors.

We analyze these results from various points of view, including the role of the time-filer, the relative performance of the descriptors and the relative performance per action.

(1) We notice that recognition performance of EHC increases with the increase of the window size of time-filter like any other descriptor. In fact, time-filter reduces the minima in the anti-diagonal direction, resulting from motion in the static descriptor. (2) The MDS is insensitive to mesh deformation which maintains the geodesic distance and shows lower recognition performances. (3) Our descriptor outperforms MDST and other classic shape descriptors (SI, SHRT, SD) and shows competitive results with (SHvrG/SHvrS) and aMRG. (4) Multiframe shape-flow matching required in SHvrG allows the descriptor to be more robust but the computational cost will increase by the size of selected time window. (5) Our EHC descriptor, by its simple representation, demonstrates a comparable recognition performance to aMRG. It is efficient as the curve extraction is instantaneous and robust as the curve representation is invariant to elastic and geometric changes thanks to the use of the elastic metric. (6) Finally, the result analysis for each action shows that TEHC gives a smooth rates that are stable and not affected by the complexity of the motion. Such complex motions are rock and roll, vogue dance, faint, shot arm as illustrated in Fig.6 (b). However, this is not the case for SHvrS where performance recognition falls suddenly with complex motions as presented in figure 18 at [11].

We apply the time filtering Extremal Human Curves descriptor to real captured 3D video sequences of people. Inter-person similarity across two people in a walking motion with an example similarity curve are shown in Fig. 7. Our temporal similarity measure identifies correctly similar frames across different people. These similar frames are located in the minima of the similarity curve.

VI. CONCLUSIONS

In this paper, a novel 3D shape descriptor for the purpose of 3D human shape similarity has been proposed. Some general rules for the extraction of extremal curves as geometric invariant descriptors of body shape within Riemannian

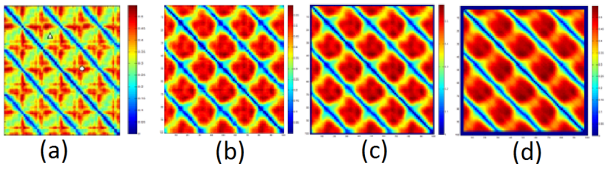

Fig. 5: Similarity measure for "Fast Walk" motion in a straight line compared with itself. Coldest colors indicate most similar frames. (a) Temporal Ground-Truth (TGT), (b-d) Self-similarity matrix computed with TEHC with window size 3, 5 and 7 respectively.

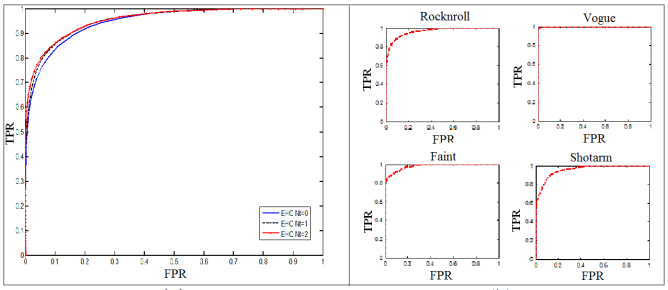

Fig. 6: ROC curves (a) for static ($N_t = 0$) and time-filtered EHC decriptor ($N_t = 1$, $N_t = 2$) on self-similarity across 14 people doing 28 motions, (b) ROC performance for 4 complex motions obtained by EHC, for fixed window size 5 ($N_t = 2$) against Temporal Ground Truth.

Shape Space framework have been discussed. Body shape in a given pose is firstly represented as a set of geodesic curves extracted from shape surface using extremal feature points. Then, an elastic metric is calculated as a pairwise descriptor distance in the Shape Space, allowing the comparison between two shape models in order to estimate their similarity. The quality of our descriptor regarding the recognition performance of pose retrieval and shape similarity in 3D video was analyzed and verified also with respect to another related recent techniques. Results obtained from extensive experiments have clearly shown the promising performance of the proposed descriptor and also the advantages of using such reduced representation of the shape model.

As for short term future work, we plan to investigate the usage of our descriptor for further related applications like 3D human action and gesture recognition.

REFERENCES

[1] K. M. Cheung, S. Baker, and T. Kanade, "Shape-from-silhouette across time part i: Theory and algorithms," *International Journal of Computer Vision*, vol. 62, no. 3, pp. 221 – 247, May 2005.

[2] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, and S. Thrun, "Performance capture from sparse multi-view video," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 98:1–98:10, Aug. 2008.

[3] T. Kanade, P. Rander, and P. J. Narayanan, "Virtualized reality: Constructing virtual worlds from real scenes," *IEEE Multimedia, Immersive Telepresence*, vol. 4, no. 1, pp. 34–47, January 1997.

[4] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3d scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 433–449, 1999.

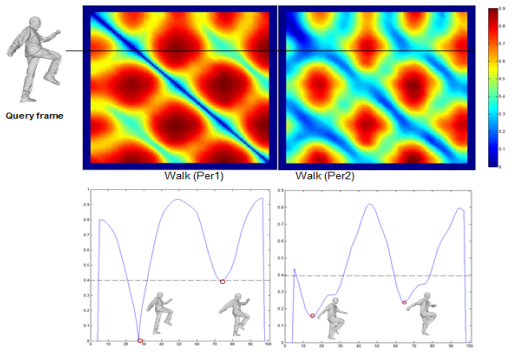

Fig. 7: Inter-person similarity measure for real data. Example of similarity matrix and similarity curve for sequences of "Walk" across 2 actors. The retrieval frames are located on the local minima of the similarity curve.

[5] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Shape distributions," *ACM Transactions on Graphics*, vol. 21, no. 4, pp. 807–832, Oct. 2002.

[6] M. Ankerst, G. Kastenmüller, H.-P. Kriegel, and T. Seidl, "3d shape histograms for similarity search and classification in spatial databases," in *Proceedings of the 6th International Symposium on Advances in Spatial Databases*, ser. SSD '99. London, UK, UK: Springer-Verlag, 1999, pp. 207–226.

[7] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3d shape descriptors," in *Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, ser. SGP '03. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2003, pp. 156–164.

[8] P. Yan, S. M. Khan, and M. Shah, "Learning 4d action feature models for arbitrary view action recognition," in *2008 IEEE Conference on Computer Vision and Pattern Recognition, 24-26 June 2008, Anchorage, Alaska, USA*, 2008.

[9] P. Huang, T. Tung, S. Nobuhara, A. Hilton, and T. Matsuyama, "Comparison of skeleton and non-skeleton shape descriptors for 3d video," in *Proceedings of the 3DPVT International Symposium*, Pairs, France, May 2010.

[10] T. Tung and T. Matsuyama, "Topology dictionary for 3d video understanding," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 8, pp. 1645 –1657, aug. 2012.

[11] P. Huang, A. Hilton, and J. Starck, "Shape similarity for 3d video sequences of people," *Int. J. Comput. Vision*, vol. 89, no. 2-3, pp. 362–381, Sep. 2010.

[12] J. Starck and A. Hilton, "Model-based multiple view reconstruction of people," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, oct. 2003, pp. 915 –922 vol.2.

[13] M. Mortara and G. Patané, "Affine-invariant skeleton of 3d shapes," in *Proceedings of the Shape Modeling International 2002*, Washington, USA, 2002, pp. 245–278.

[14] S. Katz, G. Leifman, and A. Tal, "Mesh segmentation using feature point and core extraction," *The Visual Computer*, pp. 649–658, 2005.

[15] J. Tierny, J.-P. Vandeborre, and M. Daoudi, "Invariant high level reeb graphs of 3d polygonal meshes," *IEEE 3DPVT International Symposium*, vol. 0, pp. 105–112, 2006.

[16] S. Joshi, E. Klassen, A. Srivastava, and I. Jermyn, "A novel representation for riemannian analysis of elastic curves in rn," in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, june 2007, pp. 1 –7.

[17] C. Samir, A. Srivastava, M. Daoudi, and E. Klassen, "An intrinsic framework for analysis of facial surfaces," in *Journal of International Journal of Computer Vision, Vol. (82), Number 1 / avril 2009*, 2009.

[18] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H.-P. Seidel, "A statistical model of human pose and body shape," in *Computer Graphics Forum (Proc. Eurographics 2008)*, P. Dutr'e and M. Stamminger, Eds., vol. 2, no. 28, Munich, Germany, Mar. 2009.

[19] T. Tung and F. Schmitt, "The augmented multiresolution reeb graph approach for content-based retrieval of 3d shapes," *International Journal of Shape Modeling*, pp. 91–120, 2005.